Time-Dependent Origin-Destination Estimation from Land Use and Mobile Phone Data

Suxia Gong [1*], Ismaïl Saadi [1,2], Jacques Teller [1], Mario Cools [1, 3,4]

[1] LEMA research group, Urban & Environmental Engineering Department, University of Liège, Belgium, [2] F.R.S.-FNRS, Brussels, Belgium,

[3] Faculty of Business Economics, Hasselt University, Belgium, [4] Department of Information Management, Modelling and Simulation, KU Leuven, Belgium

[*]Email: suxia.gong@uliege.be

PARTNERS

OVERVIEW

- This study utilizes mobile-phone-based OD-matrices and land use data to predict the travel demand for the province of Liège, Belgium.
- We apply the K-means clustering to the hourly aggregated net flow data to create traffic analysis zones (TAZs) clusters.
- We validate the clusters based on the land use using the logit regression model
- We deploy machine learning algorithms to estimate OD-flows considering land use features.

1. Context

Understanding the relationship between land use and dynamic human mobility is crucial to transport planning. In this study, we obtain land use data from Walloon geoportal WalOnMap.

Mobile phone data are available as two weeks' hourly aggregated OD-matrices for the province of Liège (310 zones). The calibration has been done by Wallonia SPW Mobilité et Infrastructures concerning the level of activity of users observed, Proximus market share, GSM penetration, and the number of inhabitants in Belgium.

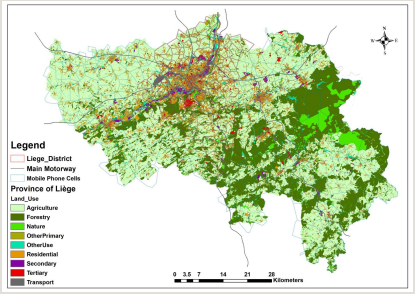

FIGURE 1 | Land use of the province of Liège

2. METHODS

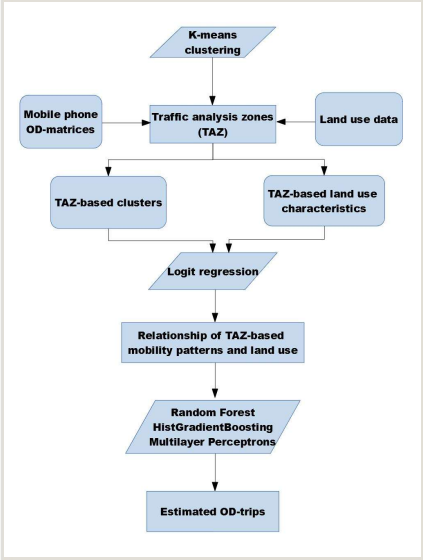

FIGURE 2 | Proposed overall modeling framework

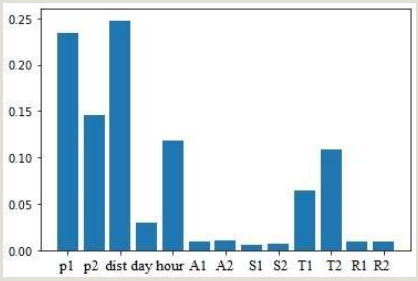

FIGURE 5 | Random Forest with feature importance score: p = population, dist = distance, day = day number, hour = 24 hour within a day, A = Agriculture percentage, S = Secondary production, T = Tertiary production, R = Residential percentage for origin (1) and destination (2), respectively

TABLE 2 | Hourly OD estimation comparison

Regressor	R^2	MSE	Accuracy
RF (without land use)	0.9871	5.2796	0.8627
RF (with land use)	0.9872	5.2720	0.8639
HGB (with land use)	0.9454	22.4833	0.7839
MLP (with land use) 30 iterations not converged	0.7331	109.9652	0.8059
MLP (with land use and spatial and temporal one-hot encoding) 30 iterations converged	0.9735	10.9131	0.8515

3. Results

a) K-means clustering

Input: normalized net flow (number of arrival minus number of departure) derived from four 24 hours OD-matrices, one regular weekday (d1), one regular weekend day (d2), one holiday weekday (d3), one holiday weekend day (d4)

Found: two clusters with Silhouette score 0.358

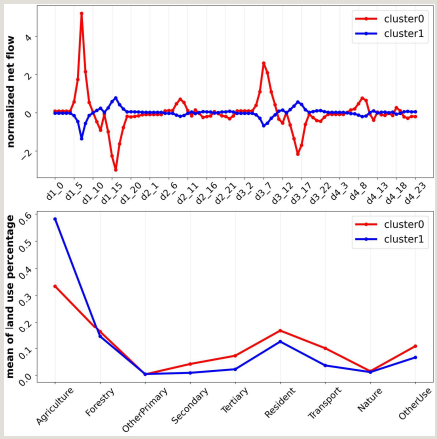

FIGURE 3 | Mean of normalized net flow over space at each time step (Upper); Mean of land use percentage for each cluster (Lower)

b) Binary logit regression

Input: Y = two clusters derived from K-means; X = land use percentage and population density for each TAZ

Found: Mcfadden R^2 = 0.298, Accuracy = 0.85,
Accuracy cluster0 = 0.64,
Accuracy cluster1 = 0.88

TABLE 1 | Logit Regression Results

	coef	std err	z	P> |z|	[0.025	0.975]
density_km2	-0.0010	0.000	-2.153	0.031	-0.002	-9.32e-05
Agriculture	2.4427	0.476	5.130	0.000	1.510	3.376
Forestry	-0.3828	0.735	-0.521	0.603	-1.824	1.058
OtherPrimary	15.1242	16.371	0.924	0.356	-16.962	47.210
Secondary	-24.7948	8.934	-2.775	0.006	-42.304	-7.285
Tertiary	-10.4264	5.402	-1.930	0.054	-21.015	0.162
Resident	10.1052	3.301	3.061	0.002	3.635	16.575
Transport	-6.9078	4.607	-1.499	0.134	-15.937	2.122
Nature	1.5578	3.863	0.403	0.687	-6.014	9.130
OtherUse	8.0946	5.382	1.504	0.133	-2.454	18.643

c) OD Estimation Liège District (123 zones)

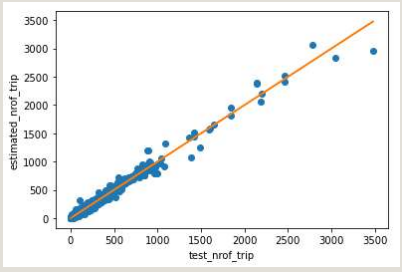

FIGURE 4 | Random Forest with land use

CONCLUSIONS

1. The framework first identifies attractive and generative TAZs based on net flow.
2. As a first validation outcome using binary logit regression, land use such as Agriculture (main primary production), Tertiary production and Residential area have significance influence on the arrival and departure trips.
3. Including land use in OD estimation based on Random Forest has slightly improved the result. However, the spatial and temporal effects need to be further studied.

REFERENCE

Gong, S., Saadi, I., Teller, J., & Cools, M. (2021). Validation of MCMC-based travel simulation framework using mobile phone data. *Frontiers in Future Transportation*, *2*, 10.
Yanyan Chen, Zheng Zhang, and Tianwen Liang. Assessing urban travel patterns: An analysis of traffic analysis zone-based mobility patterns. Sustainability, 11(19):5452, 2019.
Benjamin Beaumont, Ta¨ıs Grippa, and Moritz Lennert. A user-driven process for inspire-compliant land use database: example from wallonia, belgium. Annals of GIS, 27(2):211–224, 2021.